Hand Gesture Recognition

Akshit Maheshwari
Punjab Engineering College
Sid -19105029

Gurnoor
Punjab Engineering College
Sid-19105046

Kshitij Sethi Jain
Punjab engineering college
Sid-19105100

Ashish Ranjan
Punjab Engineering College
Sid-19105049

Aman Singla
Punjab Engineering College
Sid-19105028

Sudeep Dhara
Punjab Engineering College
Sid-19105011

Ayush Jha
Punjab Engineering College
Sid-19105051

Abstract

Pattern and gesture recognition are one of the growing fields of research. It being a significant part in non-verbal communication hand gestures are playing vital role in our daily life. Gesture recognition system supplies with an innovative, natural and user-friendly way of interaction with the machine. Gesture Recognition has a large area of application for example in human machine interaction, sign language interpreter (for visually impaired people), immersive game technology and many more. A normal human hand shape has a thumb and four fingers and by keeping in mind the picture of an alike of a human hand we will try to present a real time system for the recognition of hand gesture which will be on the basis of the training set or data which will be given in input on shape-based features which would be like orientation of hand, status of fingers (open or closed), thumb in terms of raised or folded fingers of hand and their respective location in image. The approach we tried to introduce in this research paper is being totally dependent on the shape parameters of the hand. It doesn't take any consideration of hand gesture recognition like skin colour, texture as these images-based features could be extremely variant to different light, external conditions or some other influences. To implement this algorithm, we have utilized a simple web cam of the laptop. This algorithm-based approach for hand gesture recognition can identify many different gestures on the bases of training data set we give to it. This proposed implemented algorithm has been tested very carefully and rigorously and it gives 95% of positive recognition rate.

Motivation

In human computer interaction, gesture and gesture recognition words are commonly used. A gesture is defined as the motion of the body by the user in order to convey

some meaningful information. On the other hand, gesture recognition is the process by which gesture shown by the user is identified by the system. In the view of computer vision, hand gesture is fed as an input for broad range applications. With the advancement of current user machine interaction tools and components including keyboard, joystick, mouse is not sufficient. Thus, to provide a more genuine interface to the computer system, hand gestures are used to represent different shapes which is identified by the gesture recognition system. This kind of natural interaction is the base of advance virtual environments. Keeping computers aside, if we consider human interactions, then we can realize that how much we use a wide range of gestures in our day-to-day life. In is known that people gesticulate ore while talking on phone and are not able to see face to face. Among different cultures gestures are greatly used as a mode of communication. The frequent use different kind of gestures in our everyday life as a mode of interaction motivates us to use the gestural interface and use them in a wide range of application via computer vision.

Literature survey

These are some insights of the reference we took and followed it as a bible for our project work which helped us immensely on our path during this project-:

“Hand Gesture Recognition for Human Computer Interaction”[1],- With the event of virtual environment all around us, current user-machine interaction tools and methods like mouse, joystick, keyboard and electronic pen don't seem to be enough. Hand gesture has the aptitude to represent ideas and actions very easily, thus using these different hand shapes, being identified by gesture recognition system and interpreted to get corresponding event, has the potential to produce a more natural interface to the pc system. this kind of natural interaction is that the base for virtual environments. If we ignore the globe of computers for a long time and consider interaction among kinsfolk, we will simply realize that we are utilizing a large range of gestures in our daily personal communication. By the very fact it's also shown that folks gesticulate more after they are talking on telephone and don't seem to be able to see one another as in face-to-face communication. The gestures vary greatly among cultures and context still are intimately employed in communication. the numerous uses of gestures in our way of life as a mode of interaction motivates the employment of gestural interface and employs them in big selection of application through computer vision.

“Static Hand Gesture Recognition supported Convolutional Neural Networks”[2]- during this document we got a short description of the techniques we used, our proposed methodology, and also the experiments we dispensed. the ultimate sections of this work show the results we obtained, a discussion and comparison with other works and, lastly, the conclusions and perspectives for future work.

“Hand Gesture Recognition using Convolution Neural Network ”[3]-In this paper we got more information about the

CNN and gave us the attitude why we should always prefer this method out of all the techniques and he proposed it as an algorithm for real-time hand gesture recognition . “The proposed CNN achieves a median accuracy of 95% on the dataset comprising of seven hand gestures and 301 images for every gesture” by this fact we understood the importance of using CNN.

“Deep learning in Vision Based Static Hand Gesture Recognition ”[4]-In this they proposed that applying deep learning to the matter of hand gesture recognition for the full 24 hand gestures obtained from the Thomas Moeslund's gesture recognition database. They showed that more biologically inspired and deep neural networks like convolutional neural network and stacked denoising autoencoder are capable of learning the complex hand gesture classification task with lower error rates. The considered networks are trained and tested which was supported by the info during this paper. Results comparison is then made against the earlier works during which only small subsets of the ASL hand gestures are considered for recognition.

Related Work

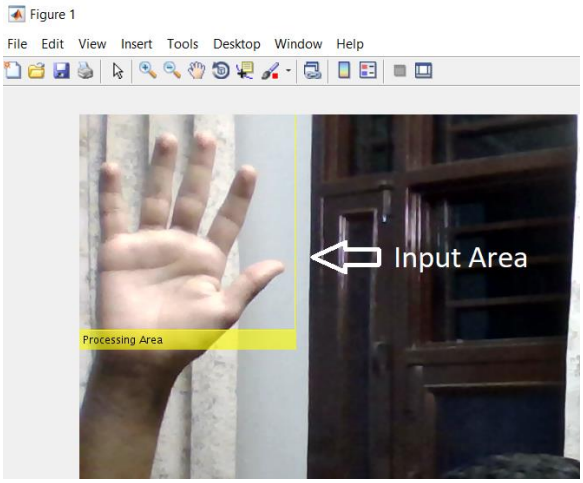
In recent times the concept of Gesture Recognition has become a very important and an influential term. Developers have invented or developed many Hand Gesture recognition techniques for determining different types of hand Gestures. Now as we know that no technique is perfect or best so, all those techniques have some pros and cons. Some of the methods are as follows, one of the oldest methods is of wired technology, in this method the users used to be tied up with wire which will lead to connection to the computer system. The disadvantage of this method is that the user will not be able to move freely anywhere or room as they are being connected by some wires. The example of wired technology is instrumented gloves which is also known as data gloves or electronic gloves. Now these gloves are used to be made or buildup of sensors and it is these sensors that gives the information regarding various hand motion or finger movements. One of the biggest drawbacks of these gloves are the are too expensive to be used. Later these gloves were replaced by optical markers. These methods were based on using infra - red light or its projection. To find the location of fingers or movement of hands these markers use infra-red light projection on screen. Now these is a quite good method and show good results but one the cons of this method is that its algorithm or its configuration is a little bit complex so it’s not that feasible. Then on the technologies keep on growing and advanced technologies came and one of them is Image based methods or technology. In these types of technologies, we need to process image features like its colour, its texture etc. As we all know that these textures and colour of person or things that vary from place to place or continents and hence the methods of these image-based recognition varies and it can be very rapid variation also. Because all these methods were not that feasible and for getting various hand gestures with as minimum variation possible, we have chosen Convolution Neural Network (CNN or Alexnet) for our gesture recognition project. These methods work on shape or figure-based recognition for gestures. In CNN we work by breaking the images into its very small components like edges and blobs. This method or algorithm involves uses of filters, contour extraction, and segmentation. The proposed or said methodology will be explained later in detail but in short, we can understand or can be visualized as, we will obtain the

images from the database, and after that image processing will began by the above-mentioned ways, like filters, contour, extraction etc.

Input

The input to the model is a segmented image containing the hand gesture.

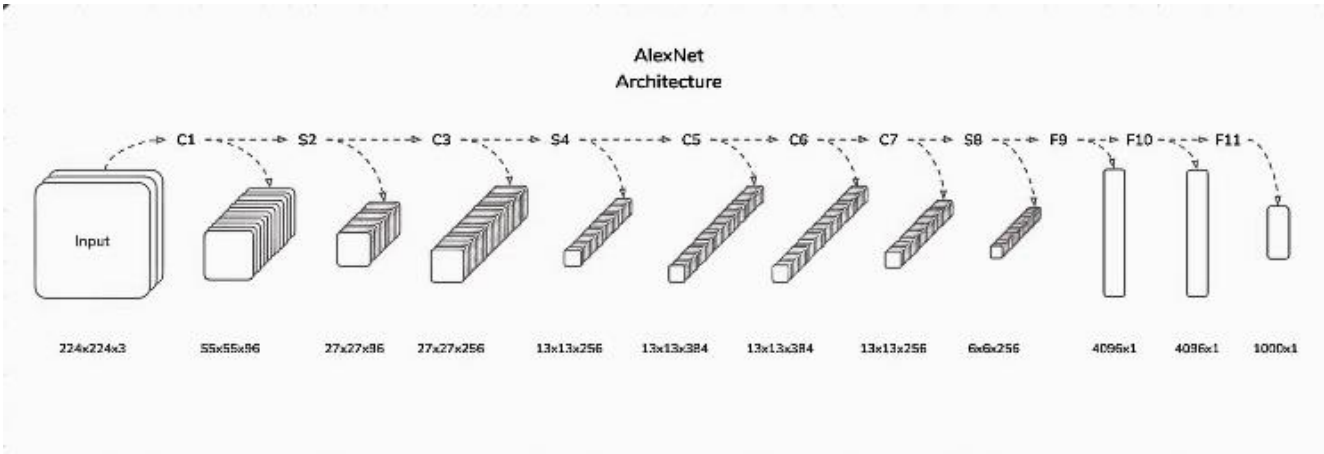
The segmentation is done manually by providing a specified space during input phase to put the hand gesture in. For training the captured images are first resized to the size of 227*227 and then manually grouped into different folders according to the hand gesture it contains.



Training

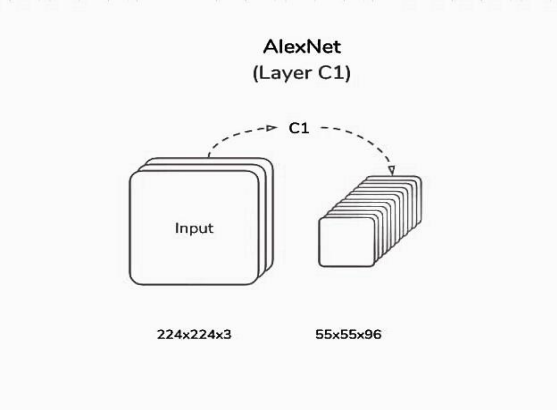
The training data is fed to Alexnet CNN for feature extraction.

A graphical representation of Alexnet CNN model is shown below.



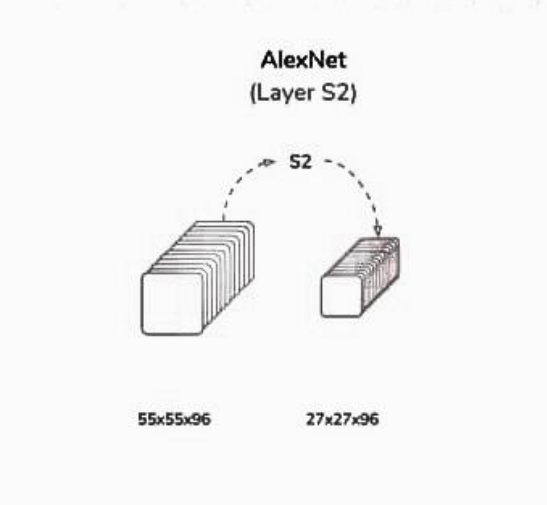
C1: First Convolution Layer

The first layer of AlexNet is a convolutional layer that take a 227*227-pixel RGB image as input. This layer performs convolution using 96(11*11) filters with a stride of 4 and padding of 2. This produces a (55 * 55 * 96) output tensor that is next passed through a ReLu activation function and then to next layer.



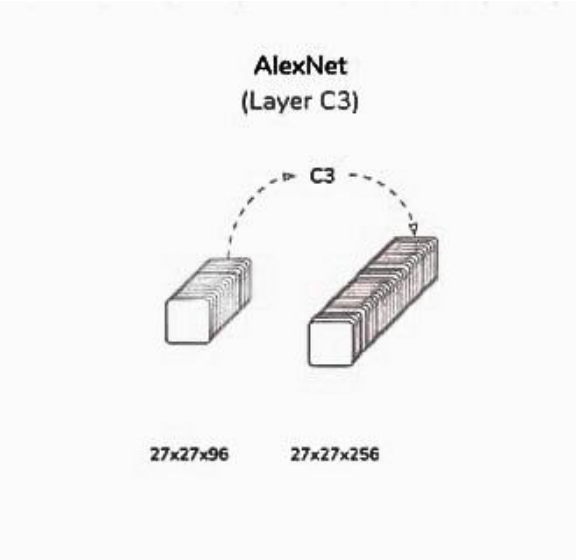
S2: First Max Pooling Layer

The 2nd layer of AlexNet is a max polling layer that takes output of layer 1 as its input. It performs a sub sampling operation using a 3*3 filter with a stride of 2. Output of this layer is a (27 * 27 * 96) tensor that is passed to next layer.



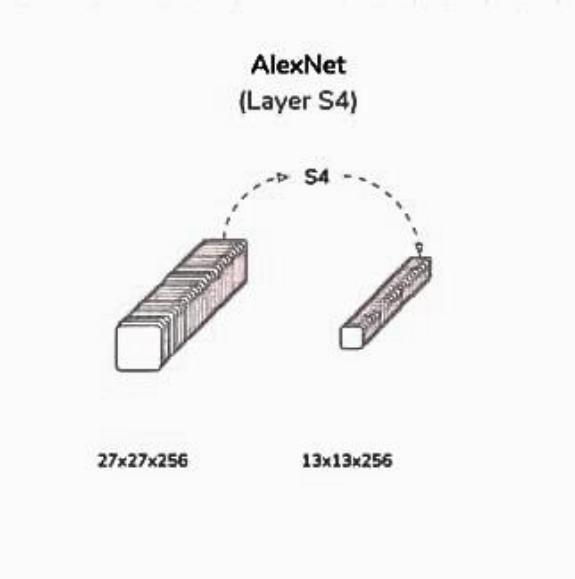
C3: Second Convolution Layer

The third layer is another convolutional layer that accepts output of layer S2 as its input. It performs convolution using 256(5*5) filters with a stride of 1. Output is a (27*27*256) tensor that is then passed through a ReLu activation function and then to next layer.



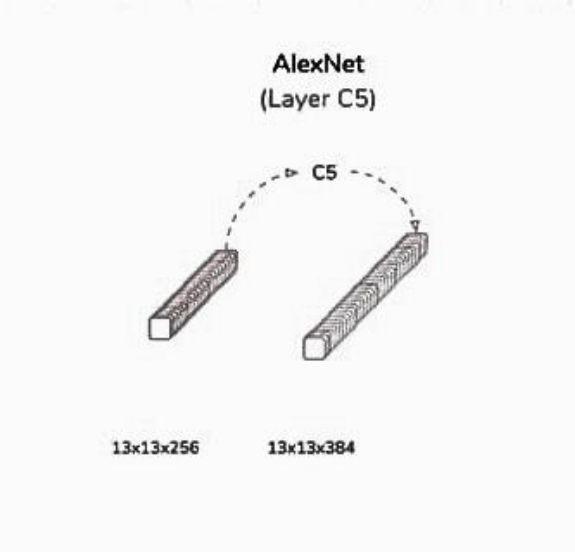
S4: Second Max Pooling Layer

Fourth layer is a max pooling layer whose input is output of 3rd layer. It performs sub sampling using (3*3) kernel with stride of 2 and produces a (13 * 13 * 256) tensor as output which is passed through a ReLu activation function and then to next layer.



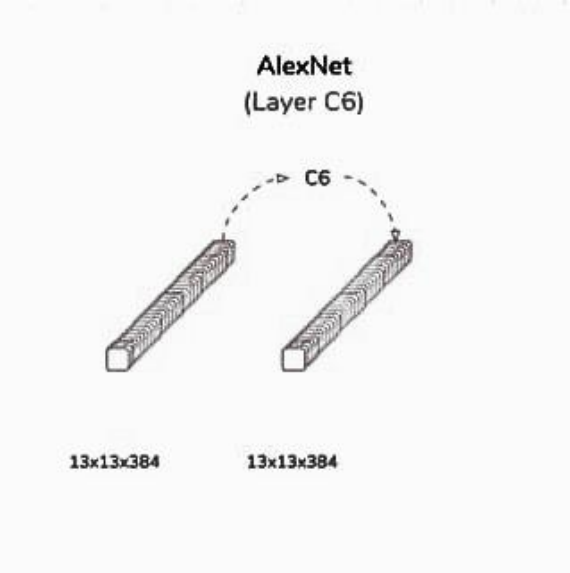
C5: Third Convolution Layer

Fifth layer is another convolutional layer. It performs convolution using 384 (3*3) filters with a stride of 1 and produces (13*13*384) tensor as output which is then passed to a ReLu activation function and then to next layer.



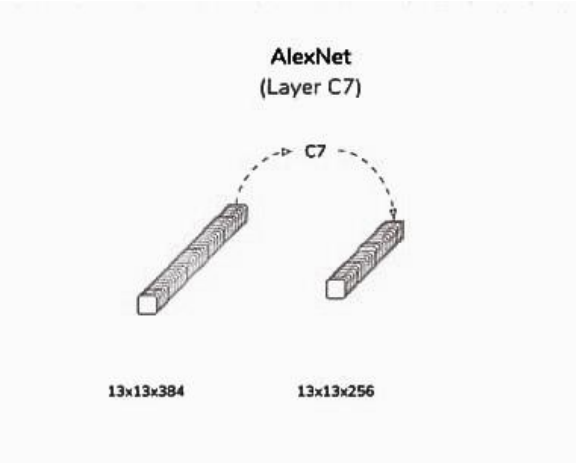
C6: Fourth Convolution Layer

Sixth layer is a convolutional layer that performs operation similar to fifth layer, thus producing output of similar size which is then passed to next layer.



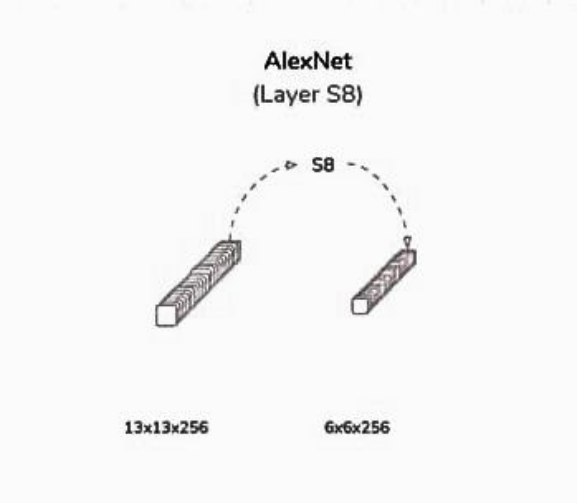
C7: Fifth Convolution Layer

Seventh layer is also a convolutional layer that performs convolution using 256(3*3) filters with a stride of 1 and produces (13*13*256) tensor as output. Output is then passed through a ReLu activation function and then to next layer.



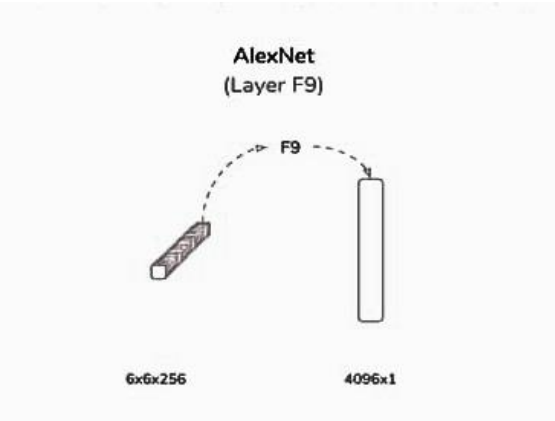
S8: Third Max Pooling Layer

The eighth layer is a max pooling layer that performs sub sampling operation using a (3*3) window region with a stride of two. This produces (6*6*256) tensor as output which is then passed to next layer.



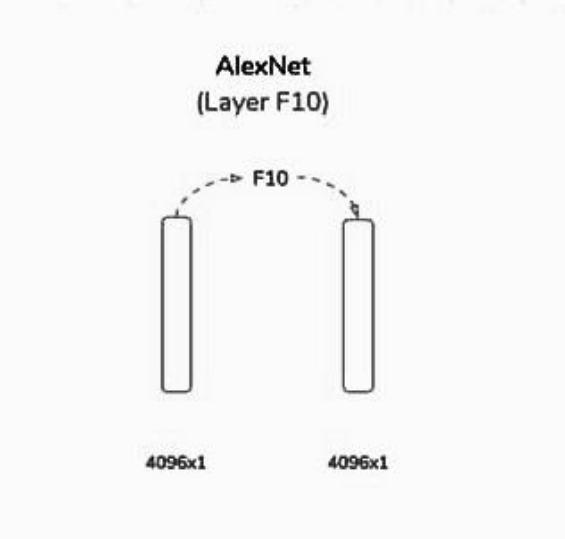
F9: First Fully Connected Layer

The ninth layer is a fully connected layer that performs weighted sum operation with an added bias term and produces (4096*1) tensor as output which is then passed through a ReLu activation function and then to next layer.



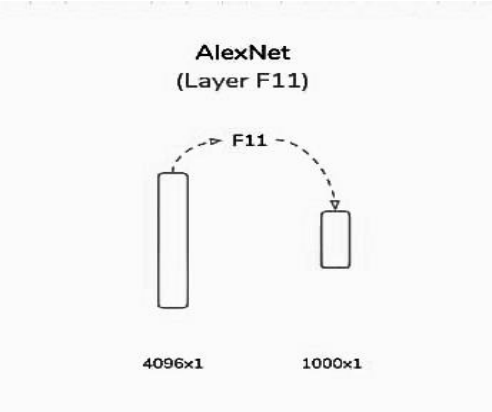
F10: Second Fully Connected Layer

The tenth layer is another fully connected layer that performs same operation as F9 and produces same (4096*1) tensor as output which is then passed through a ReLu function and then to next layer



F11: Third Fully Connected Layer

The eleventh and the final layer is also a fully connected layer that performs same operation as F9 and F10 and produces (1000*1) tensor as output that is then passed through a softmax activation function. The output of softmax activation function contains prediction of the network.

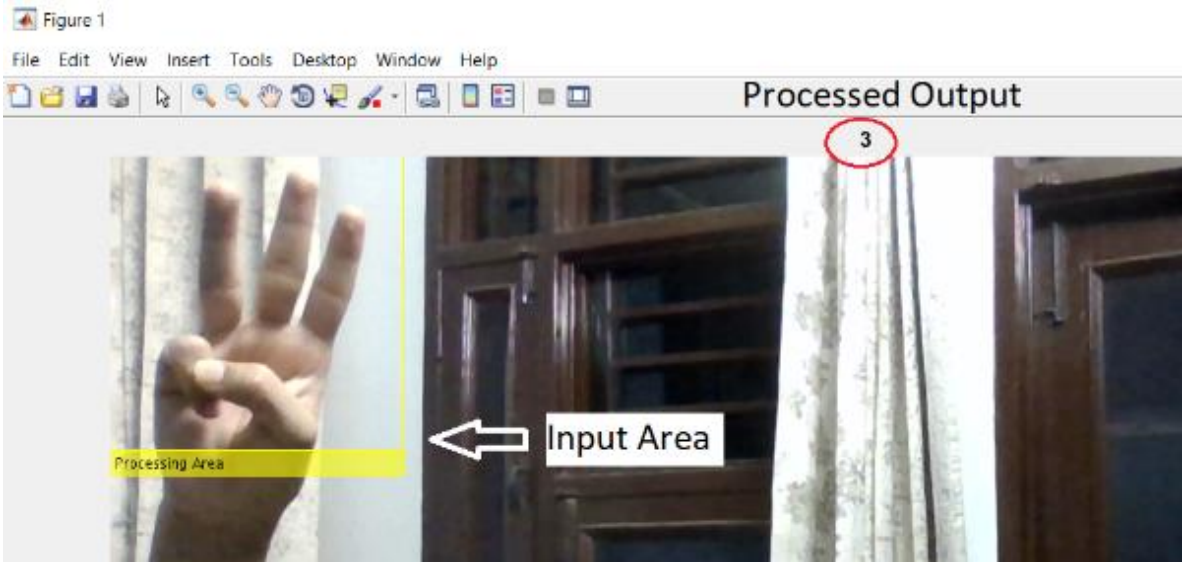


Testing

For testing the input is taken from camera.

Then the captured image is resized to 227*227 pixels and then passed to the model for processing.

The model then labels the image with the group name it belongs to.



Reference: -

1) Meenakshi Panwar and Pawan singh Mehra's document - "Hand Gesture Recognition for Human Computer Interaction", in proceedings of IEEE International Conference for Image Processing , Wakanghat , India, November 2011.

2) Raimundo F. pinto, Carlos D.B. Borges , Antonio almeida , "Static Hand Gesture Recognition for Convolutional Neural Networks",

3) Felix Zhan's document, "Hand Gesture Recognition using Convolution Neural Network".

4) Oyebade k. Oyedotum, and Adnan Khashman's research on- "Deep learning in Vision Based Static Hand Gesture Recognition".

5) Xiaohui Shen, Lance williams Ying yu's- "Dynamic hand gesture recognition: An exemplar-based approach from motion divergence fields".